Estimation of Hourly Link Population and Flow Directions from Mobile CDR

Ko Ko Lwin 1,*, Yoshihide Sekimoto 1 and Wataru Takeuchi 2

1 Human Centered Urban Informatics, Institute of Industrial Science, The University of Tokyo, Tokyo 153-8505, Japan; sekimoto@iis.u-tokyo.ac.jp
2 Remote Sensing of Environment and Disaster, Institute of Industrial Science, The University of Tokyo, Tokyo 153-8505, Japan; wataru@iis.u-tokyo.ac.jp
* Correspondence: kokolwin@iis.u-tokyo.ac.jp

Received: 27 September 2018; Accepted: 14 November 2018; Published: 17 November 2018

Abstract: The rise in big data applications in urban planning and transport management is now widening and becoming a part of local government decision-making processes. Understanding people flow inside the city helps urban and transport planners build a healthy and lively city. Many flow maps are based on origin-and-destination points with crossing lines, which reduce the map’s readability and overall appearance. Today, with the emergence of geolocation-enabled handheld devices with wireless communication and networking capabilities, human mobility and the resulting events can be captured and stored as text-based geospatial big data. In this paper, we used one-week mobile-call-detail records (CDR) and a GIS road network model to estimate hourly link population and flow directions, based on mobile-call activities of origin–destination pairs with a shortest-path analysis for the whole city. Moreover, to gain the actual population size from the number of mobile-call users, we introduced a home-based magnification factor (h-MF) by integrating with the national census. Therefore, the final output link data have both magnitude (actual population) and flow direction at one-hour intervals between 06:00 and 21:00. The hourly link population and flow direction dataset are intended to optimize bus routes, solve traffic congestion problems, and enhance disaster and emergency preparedness.

Keywords: big data; CDR origin–destination pairs; shortest-path analysis; hourly link population and flow direction; home-based magnification factor (h-MF)

1. Introduction

Information on human mobility (both magnitude and direction) inside the city is important in urban and transport planning, such as bus route planning and optimization [1], trip frequency scheduling [2], transportation modes prediction [3], traffic congestion management [4], public facility management, and disaster and emergency preparedness [5], in order to build a healthy and lively city. Today, big data applications in urban planning and transport management are widening and are becoming part of local government decision-making processes. Traditional ways of acquiring people flow inside the city are paper-based travel surveys or other transport statistics, which are expensive and labor intensive. Today, human mobility and activities can be tracked using mobile phone call activities (call-detail record, CDR), internet usage, and other social interacting events through online social networks (OSN) and other wireless sensor networks (WSNs). Although CDR has some limitation on data acquisition because of privacy protection, CDR is one means to identify the mass movement of people inside a city or across the country inexpensively and time-effectively. CDR data have been used by many researchers for origin–destination trip generation [6–11], travel behavior analysis [12–14], social interaction [15,16], urban analysis [17], and population estimation [18,19].
Generally, traffic-count information is acquired from observed data, such as traffic-count surveys, electronic toll collection (ETC), and gates and other roadside sensors. Many researchers attempt to estimate link flow from observed data [20], trip questionnaire surveys, and traffic-count survey data [21], and generate origin-and-destination (OD) matrices from automatic vehicle identification data [22]. Moreover, CDR data have been used for road usage and link traffic volume, and OD flow estimation [23–25]. In general, traffic flow measurements include not only vehicles but also travelers by measurement of their speed, direction, and magnitude at specific intersections and in specific lanes.

From a cartographic point of view in GIS, many flow-mapping algorithms and methods have been developed for visualization of either location-based flow maps (i.e., migration flow, commodity flow, flight paths, etc.) or non-spatial flow maps (i.e., Sankey diagrams, Chord diagrams, thermal flow diagrams, system flow diagrams, etc.) [26–28]. Flow maps can be represented as many-to-many (many origins to many destinations) or one-to-many (one origin to many destinations) in visualization methods. Many flow-mapping approaches use OD or source and target locations, which do not convey detailed movement patterns along the flow path. Moreover, many flow maps have some problems in readability and overall appearance because of crossing lines on the map. Within the scope of interactive mapping (i.e., digital cartography or Web-GIS), many paper-based cartographic problems can be overcome by means of setting visibility map scales of individual features and symbols or by adjusting the placement and orientation of labels by preventing them from overlapping with each other.

In this study, we estimate the link population (magnitude) and direction from one-week of CDR data to map people-flow patterns for the whole city at one-hour intervals. The main objective of the paper is to use mobile CDR data to estimate hourly link population and their flow direction at a regional scale to use in future urban and transportation planning processes. This paper is organized as follows. First, we list the data used in this study and describe the study area. Second, we explain our data-handling steps for CDR data preprocessing, home-based magnification factor computing, and generation of hourly link population and flow directions. Finally, we discuss the results and validation procedure.

2. Study Area

The study area is greater Yangon city, a major business center of Myanmar located in Southeast Asia. Every day, approximately six million commuters are moving in the city. In 2014, the population of Yangon Division was 7,360,703 and the urban population was 5,160,512 [29] (Figure 1). Yangon city has a major development plan to improve the current public bus system and many housing projects, including new urban mass rapid transit lines into central Yangon [30].
Figure 1. Yangon city road network patterns (left) and census population in 2014 (right).

3. List of Data

Myanma Posts and Telecommunications’ (MPT) call-detail record data were used in this study, along with census and road network data. Table 1 shows the list of data, sources, attribute information, and the purpose of its use in the study.

Table 1. List of data, source, attribute information, and purposes. OD—origin-and-destination.

<table>
<thead>
<tr>
<th>Dataset Name</th>
<th>Data Source</th>
<th>Acquisition Date</th>
<th>Attribute Information</th>
<th>Purpose</th>
</tr>
</thead>
<tbody>
<tr>
<td>MPT CDR</td>
<td>Myanmar Posts and Telecommunications (MPT) call-detail records (CDR) for Myanmar</td>
<td>1 December 2015 to 7 December 2015</td>
<td>CDR for both voice and data, including text messages Attributes:</td>
<td>To generate OD pairs for individuals</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>• Subscriber ID</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>• Event ID</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>• Site ID</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>• Start Time</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>• Duration</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>• Download</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>• Upload</td>
<td></td>
</tr>
<tr>
<td>GIS road network</td>
<td>Yangon City Development Committee (YCDC)</td>
<td>2013</td>
<td>Road names, type, length, etc.</td>
<td>To compute trip distance, speed, and direction</td>
</tr>
<tr>
<td>National census</td>
<td>Myanmar Information Management Unit (MIMU)</td>
<td>2014</td>
<td>Demographic data by township, state, and division. Attributes:</td>
<td>To compute home-based magnification factor (h-MF)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>• male, female, and total</td>
<td></td>
</tr>
<tr>
<td>JICA Transport Surveys Report</td>
<td>Japan International Cooperation Agency (JICA)</td>
<td>Transport surveys conducted between February 2013 and August 2013</td>
<td>Traffic volume of major roads by vehicle type Number of passengers of public transport (bus, ferry, and railway) Trip patterns by transport modes Number of users by parking</td>
<td>Result validation</td>
</tr>
</tbody>
</table>
4. Methodology

4.1. Research Flow

In this research, we used one-week MPT mobile CDR data to extract OD trips at one-hour intervals from 06:00 to 21:00. Later, these OD pairs were used to compute individual travel distance, duration, speed, and direction, based on GIS road network data models. Finally, we used these data to generate link population and flow directions for each road link (edge) by aggregating the total number of mobile users with their magnification factors. In this research, CDR data processing has been done by BigGIS-RTX research toolbox [31]. Figure 2 gives an overview of this research.

![Data processing and research flow](image)

**Figure 2.** Data processing and research flow. MPT—Myanmar Posts and Telecommunications; CDR—call-detail records; OD—origin-and-destination.

4.2. MPT CDR Data Preprocessing

We used the anonymized subscriber ID from mobile CDR data obtained from MPT for a seven-day period (1 December 2015–7 December 2015), for the whole country. MPT owned 46% of the market share in 2015 [32], along with Telenor, Ooredoo, and other providers. MPT had 20 million subscribers in May 2016 [33], which represents 40% of the national population. MPT is a state-owned enterprise in Myanmar, and is under the supervision of the Ministry of Transport and Communications; it operates a nationwide network infrastructure, with the widest 3G mobile network coverage throughout Myanmar,
recorded at 95% in March 2016. This includes both voice and data (Figure 3) with encrypted SIM card ID, Event Code, Call Time, Call Duration, Upload Data Size, and Download Data Size. Because of the nature of the big data collection system, preprocessing steps are required to use these data in this study, such as removing empty call durations and records; formatting the strings; and converting some numerical values, such as site ID, into strings to avoid manipulation during the process. In this dataset, the number of voice users was larger than that of data users.

![Formatted MPT CDR for both voice and data with base transceiver station (BTS) locations.](image)

**Figure 3.** Formatted MPT CDR for both voice and data with base transceiver station (BTS) locations.

### 4.3. Home User-Based CDR Data Magnification Factor

Here, we introduced the home-based magnification factor (h-MF) to obtain the actual population from mobile users. The computation of h-MF has three steps. First, find each person’s home location Cell-ID and sum the total persons by Cell-ID called Cell-ID home users. Second, find the Cell-ID disaggregated census population based on Cell-ID home users. Third, compute the Cell-ID magnification factor and assign to individual persons (person ID (PID)) who belong to this Cell-ID.

First, we found each person’s home location Cell-ID, which was extracted from seven-day CDR data calls between 20:00 and 07:00, based on the assumption that users stay at home (Figure 4a). After that, we found each person’s maximum call frequency of Cell-ID (Figure 4b) and assumed that this Cell-ID was the person’s home location. In this step, we obtained Cell-ID total home users and its belonging persons (PIDs; Figure 4c).

Second, we found the Cell-ID disaggregated census population, based on census population and Cell-ID home users obtained from the previous step. Generally, the census population was available by township or block with the aggregated population. We wanted to disaggregate census population into the Cell-ID population. Equation (1) (Figure 4d) was used to disaggregate census population, based on Cell-ID home users. This equation was modified from our previous study, building population estimation in GIS [34].

Third, we computed individual Cell-ID h-MF using Equation (2). However, one-week CDR data could not obtain all home users in the dataset, known as unknown home users. Some subscribers did not make any calls between 20:00 and 07:00 within a seven-day period. In this case, we used the default value, which was calculated by subtracting the total magnified population from the country population and dividing by unknown home users (Equation (3) in Figure 4). We found that 20% of the PIDs should be assigned a default magnification factor for the whole country. Finally, Cell-ID h-MF was assigned to its home users acquired in the first step. Therefore, the MF is fully synchronized with
human mobility; wherever an individual goes, their MF will follow them. The total population can be obtained by summing all MFs in an area (Figure 4e).

**Figure 4.** Graphical illustration of home-based magnification factor, link counts, population, and flow direction computational steps. (a) Home user extraction; (b) Finding PID with maximum call frequencies; (c) Counting total home users by Cell-ID; (d) Disaggregation of Cell-ID population with its home users; (e) Illustration of link counts, population and flow directions computation.

### 4.4. Hourly Link Population and Flow Directions

The following steps were performed to achieve link population and flow directions.

Moving cell tower locations to the nearest road nodes: If the road link was too long, curved, or twisted, measurement of direction could vary depending on the location of the start and endpoints on the link
To start every trip at the nearest road node, we moved all Cell-ID locations to the nearest road nodes (Figure 5b). Therefore, each road link could have only two opposite directional values, such as east or west, north or south, northeast or southwest, and southeast or northwest. Later, we grouped them into two categories, based on these two opposite directions, to determine the majority and minority flow directions.

**Extract hourly OD pairs:** After preprocessing was completed, we started to extract OD pairs for individual persons by pairing their successive calls or data usages (Figure 5c). During this process, we also computed several trips, duration, distance, and speed of each pair for further travel behavior and mode-choice analysis. In this process, we omitted the points where the user stayed at the same location (if successive calls or data usages are the same Cell-ID or coordinates), also known as stay points.

![Figure 5. Moving cell tower location to nearest road node to synchronize the directions and shortest path analysis. (a) Before moving Cell-ID; (b) After moving Cell-ID; (c) find the shortest path between two successive calls.](image)

**Compute link population and directions:** In this step, we used a GIS road network data model to compute route paths and directions between each OD pair (i.e., start Cell-ID x, y; and end Cell-ID x, y), based on shortest-path analysis using the well-known modified Dijstra algorithm. Direction was computed between two start and end nodes by the angle between $0^\circ$ and $360^\circ$, and grouped into eight categories, namely, north, northeast, east, southeast, south, southwest, west, and northwest. We counted unique PIDs by each road link and summed their magnification factor (h-MF) for link population to gain the actual population from mobile users (Figure 4e). Link population and flow directions were grouped into two opposite directions (i.e., east or west, north or south, northeast or southwest, etc.) to determine the majority flow direction at specific time intervals.

5. Results and Discussion

The final goal of this research was to produce hourly link population and flow direction maps to use in various urban and transport planning projects. Figures 6 and 7 show the flow magnitude by line thickness (i.e., number of people) by each road link in the morning, 06:00–07:00, and evening,
17:00–18:00. According to the map, in the morning, high magnitude links were found in many rural areas where people are entering the central business area (Figure 6), and in the evening, high magnitude links were found in downtown areas, especially bus station centers, where people are departing for various regions (Figure 7).

We can also visualize the link flow directions for both roadsides (Figures 8 and 9). The arrow direction is the majority people-flow direction from two opposite directions. For example, if the road link has 369 southbound travelers and 287 northbound, the direction arrow will be headed south (Figure 10a). The hourly link people-flow dataset can generate an hourly profile line with estimated population number for both directions at specific points on the map (Figure 10b). Later, we will build a Web-GIS for geovisualization by providing interactive mapping and spatial decision-making functions for the city and urban transport planners, policy makers, and other spatial information users.

Result validation for dynamic population estimation is a considerable challenge for many researchers, because human mobility changes over space and time. Because there is no single population estimation model to predict an accurate population in a quantitative approach, here, we used a qualitative approach (visual interpretation) rather than a quantitative approach. The result was validated by counting moving vehicles in each direction at specific points (Figure 10c). We also used a web-based, real-time smartphone geospatial data collection system [35] to collect some passenger counts at every bus stop on selected routes to measure people-flow patterns along the bus route. The people-flow magnitude and direction of this study is directly proportional to the traffic volume and directional data, which we collected at the ground. Moreover, transport surveys conducted in 2013 by Japan International Cooperation Agency (JICA) data also helped to validate the results by comparing traffic volume with link population in specific intersections and lanes.

Figure 6. Link population (magnitude) between 06:00 and 07:00.
**Figure 7.** Link population (magnitude) between 17:00 and 18:00.

**Figure 8.** Hourly link flow direction between 06:00 and 07:00.
Figure 9. Hourly link flow direction between 17:00 and 18:00.

Figure 10. (a) Hourly link flow direction with population values for both directions between 06:00 and 07:00. (b) Comparison of hourly flow magnitude for both southbound and northbound at Hledan Junction. (c) Ground-truth data collection for result validation.
6. Conclusions

We used one-week mobile CDRs to generate hourly OD pairs of individuals and found the shortest paths between the pairs. Here, we introduced home-based magnification factors to estimate actual population scaling by number of mobile home users and disaggregated census population of each base transceiver station (BTS). We computed the link population and flow directions for the whole city, which allowed visualizing detailed movement of people inside the city at one-hour intervals. Moreover, using Web-GIS, we delivered link population and flow-direction information to online geospatial information users to perform additional tasks, such as generating hourly population profile lines at specific points or links and identifying average population by user-defined areas, as well as other digital cartographic functions in their decision-making processes. Owing to some limitations in researching with CDR data, vehicle- or people-flow mapping for the whole city was limited at the time and required additional ground-truth data collection and an advanced wireless monitoring system to predict these for the whole city. Although CDR has some limitations in human mobility studies (for example, mobility depends fully on their call and data usage activities, actual traveling paths may differ from shortest paths in GIS analysis, and cell tower locational information is sometimes inaccurate owing to load balancing), CDR remains one of the data sources for a massive number of human mobility patterns inside the city or across the country. We hope that hourly link population and flow direction along with other information generated from this study, such as the number of trips per person, average travel distance, duration, speed, and mode choice, will be used in the Yangon City Development Planning process in the near future.

Author Contributions: All authors contributed substantially to this work.

Funding: Japan Science and Technology Agency and the Japan International Cooperation Agency.

Acknowledgments: We thank Myanma Posts and Telecommunications for providing the one-week CDR data for this study. We also acknowledge the Japan Science and Technology Agency and the Japan International Cooperation Agency for funding and implementing the project entitled “Development of a Comprehensive Disaster Resilience System and Collaboration Platform in Myanmar”, led by Kimiro Meguro, International Center for Urban Safety Engineering at the Institute of Industrial Science, University of Tokyo, under the Science and Technology Research Partnership for Sustainable Development (SATREPS) program.

Conflicts of Interest: The authors declare no conflict of interest.

References


